CAPÍTULO II: Distribuciones Muestrales de Probabilidad.

→ DISTRIBUCIÓN MUESTRAL DE LA MEDIA.

- $\bar{X} = \dfrac{(X_1 + X_2 + \ldots + X_n)}{n}$

X_1: se refiere a la 1ª observación;

$X_1 \to N(\mu_x; \sigma_x^2)$

⋮ (así con todas las "x").

- $E[\bar{X}] = \mu_{\bar{x}} = \mu_x$ El valor esperado de la distribución muestral de medias es SIEMPRE igual a la media poblacional.

- $Var(\bar{X}) = Var \dfrac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \dfrac{\sigma_x^2}{n} \left(\dfrac{N-n}{N-1} \right)^* = \sigma_{\bar{x}}^2$

* muestreo sin reemplazamiento.

Si $n \geq 30$ → DISTRIBUCIÓN NORMAL.

- $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{\Sigma(X_i - \mu_x)}{N}}$

A medida que el tamaño de la muestra aumenta, la varianza disminuye. Cuanto menor es el tamaño de la muestra, la varianza de la distribución muestral de medias será mayor.

→ DISTRIBUCIÓN MUESTRAL DE LA VARIANZA:

"La varianza muestral mide la dispersión entre las observaciones." (S

CONCEPTOS PREVIOS

a) Dado un conjunto de n variables Z, de naturaleza normal con valor esperado 0 y con varianza 1.:

Si $Z_i \sim N(0;1)$,

$Z_1^2 + Z_2^2 + \dots + Z_n^2 = \chi_n^2$ (Chi-cuadrado con "n" grados de libertad).

$\left(Z = \frac{X_i - \mu_x}{\sigma_x} \right)$

Si $X_i \sim N(\mu_x; \sigma_x^2)$,

$$\sum_{i=1}^{n} \left(\frac{x_i - \mu_x}{\sigma_x} \right)^2 = \chi_n^2$$

Además, $E(\chi_n^2) = n$.

$Var(\chi_n^2) = 2n$.

b) Dadas dos v.a. independientes que siguen una distribución Chi-cuadrado, la suma de éstas será otra v.a. con distribución Chi-cuadrado, con grados de libertad igual a la suma de los grados de libertad de las dos v.a. independientes anteriores.

- $S^2 = \frac{\sum_{n=1}^{i} (x_i - \bar{x})^2}{n-1}$

* La varianza no puede ser negativa, y por lo tanto siempre será positiva, de ahí que no pueda seguir una distribución normal.

- $\frac{(n-1) S^2}{\sigma^2} = \chi^2_{n-1}$

- $E(S^2) = \sigma^2$

- $Var(S^2) = \frac{2\sigma^4}{n-1}$.

→ DISTRIBUCIÓN MUESTRAL DE LA MEDIA (σ_x^2 desconocida)

$$X \to N(\mu_x; \sigma_x^2) \; ; \; \bar{x} \to N\left(\mu_{\bar{x}} = \mu_x ; \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}\right)$$

Tipificando:

$$\frac{\bar{x} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} \simeq z$$

Si σ_x^2 ¿?, entonces $\hat{\sigma}_x^2 = S^2$,

$$\frac{\bar{x} - \mu_x}{\frac{S}{\sqrt{n}}} \simeq t_{n-1}$$

* si se da el MSR, añadir el factor corrector en el denominador.

"La "t Student" la utilizaremos para llevar a cabo inferencias sobre la media de la población cuando la varianza poblacional sea desconocida".

→ DISTRIBUCIÓN MUESTRAL DE LA VARIANZA: análisis que involucran a dos poblaciones.

* Distribución F Fisher:

$$F_{v_1,v_2} = \frac{\frac{\chi^2_{v_1}}{v_1}}{\frac{\chi^2_{v_2}}{v_2}}$$

Dominio de F_{v_1,v_2} es $(0, +\infty)$

$$F_{v_1;v_2;\alpha} = \frac{1}{F_{v_2;v_1;1-\alpha}}$$

Extraemos sendas muestras de ambas poblaciones X e Y con distribución normal (tamaño de las muestras: "n_x" y "n_y").

$X \to N(\mu_x; \sigma_x^2)$ $\qquad$ $Y \to N(\mu_y; \sigma_y^2)$

$$\frac{(n-1) S_x^2}{\sigma_x^2} = \chi^2_{n_x-1} \qquad \frac{(n-1) S_y^2}{\sigma_y^2} = \chi^2_{n_y-1}$$

$$\frac{\dfrac{\chi^2_{n-1}}{n_x-1}}{\dfrac{\chi^2_{n_y-1}}{n_y-1}} = F_{n_x-1;\, n_y-1} = \frac{\dfrac{S_x^2}{\sigma_x^2}}{\dfrac{S_y^2}{\sigma_y^2}}$$

Expresión utilizada para saber si las varianzas de las dos variables iniciales de partida son iguales. En el caso en que sean iguales, esto es, homocedásticas, la expresión anterior queda reducida a:

$$\text{Si } \sigma_x^2 = \sigma_y^2 \Rightarrow \frac{S_x^2}{S_y^2} = F_{n_x-1;\, n_y-1}.$$

→ DISTRIBUCIÓN MUESTRAL DE LA MEDIA: análisis que involucra dos poblaciones.

Partimos de dos v.a "x" e "y" independientes y normales. Así:

$x \to N(\mu_x; \sigma_x^2)$ $\qquad$ $y \to N(\mu_x; \sigma_y^2)$

⇓ de donde

$\bar{x} \to N\left(\mu_x; \dfrac{\sigma_x^2}{n_x}\right)$ $\qquad$ $\bar{y} \to N\left(\mu_y; \dfrac{\sigma_y^2}{n_y}\right)$

$\bar{x} - \bar{y} \to$ Naturaleza normal.

$\mu_{\bar{x}-\bar{y}} = \mu_x - \mu_y.$

$\sigma^2_{\bar{x}-\bar{y}} = \dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}$

$\Bigg\}\ \bar{x}-\bar{y} \to N\left(\mu_x - \mu_y; \dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}\right)$

1er Caso: Varianzas poblacionales conocidas pero distintas.

Tipificando la distribución N. de la diferencia de medias:

$$\frac{\bar{X}-\bar{Y}-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} = z$$

2º Caso: Varianzas poblacionales conocidas e iguales.

$$\frac{\bar{X}-\bar{Y}-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}}} = z.$$

* σ^2: varianza común.

3er Caso: Varianzas desconocidas, pero por la prueba de homocedasticidad decimos que son muy iguales.

$$t_{n_x+n_y-2} = \frac{\bar{X}-\bar{Y}-(\mu_x-\mu_y)}{\sqrt{\frac{S_p^2}{n_x}+\frac{S_p^2}{n_y}}}$$

* S_p^2: varianza combinada.

$$S_p^2 = \frac{(n_x-1)S_x^2+(n_y-1)S_y^2}{n_x+n_y-2}$$

4º Caso: Variables desconocidas y distintas según la prueba de homocedasticidad.

$$\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}}=t$$

* grados de libertad: $\frac{(n_x-1)(n_y-1)}{(n_y-1)c^2+(1-c)^2(n_x-1)}$ * $c=\frac{\frac{S_x^2}{n_x}}{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}$.

Si los grados de libertad no son un número entero, se redondean por defecto.

* Si los grados de libertad son suficientes, aproximamos $t \to z$.

→ DISTRIBUCIONES MUESTRALES DERIVADAS DE LA BINOMIAL.

- La distribución muestral del número de éxitos.

De la población inicial conocemos el parámetro "p": porcentaje o probabilidad de aparición de la modalidad éxito en la población. Conocemos tb. "q": probabilidad o porcentaje de fracasos. "$(1-p)=q$". Extraemos todas las muestras posibles de tamaño "n" y definimos "r" como "nº de éxitos en una muestra de tamaño n".

Los valores "r" de cada una de esas muestras constituyen la DISTRIBUCION MUESTRAL DE EXITOS.

2) MSR → naturaleza de esta distribución es hipergeométrica:

$E(r) = n.P \qquad Var(r) = nPQ\left(\frac{N-n}{N-1}\right)$

Nat. (r) → Hipergeométrica.

Discreta: puede llegar a tomar (n+1) valores. (incluye el "0").

La probabilidad de que "r" tome un determinado valor "k" será:

$$P[r=k] = \frac{\binom{NP}{k}\binom{NQ}{n-k}}{\binom{N}{n}}$$

Si los valores de "N" y/o "n" son suficiente/ grandes, podemos acercar el modelo hipergeométrico al binomial.

Para calcular la probabilidad de un rango muy amplio de valores de "r", aproximamos la distribución binomial a una continua. Para poder llevar a cabo esta aproximación debe cumplirse que el producto nPQ sea al menos igual a 3.

Por haber ajustado una distribución de probabilidad discreta a una continua, debemos aplicar la corrección de continuidad, que consiste en la ±

de media unidad de medida a los valores de la distribución de probabilidad continua.

b) MCR → naturaleza de esta distribución es binomial. (posible aproximación a la normal cuando $nPQ \geq 3$).

$E(r) = n.P$ $\quad Var(r) = nPQ$.

Nat. (r) → Binomial.

Discreta: puede tomar (n+1) valores.

$$P(r=k) = \binom{n}{k} P^k Q^{n-k}.$$

Si $nPQ \geq 3$ entonces:

$r \to X$

Discreta → Continua.

Binomial → Normal.

→ DISTRIBUCION MUESTRAL DEL NUMERO DE EXITOS.

$\bar{p}$: "proporción de éxitos en la muestra de tamaño n."

$$\bar{p} = \frac{r}{n} = \frac{\text{Nº de éxitos}}{\text{Tamaño de la muestra}}.$$

$E(\bar{p}) = P$.

$$Var(\bar{p}) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$

$$Nad.(\bar{p}) = Nad(r).$$

→ DISTRIBUCIONES DERIVADAS DE LA BINOMIAL: análisis que involucran dos poblaciones.